$X_1, \dots, X_n \sim Be(p_X)$

$Y_1, \dots, Y_n \sim Be(p_Y)$

iid. (Xs & Ys are independent)

$H_0: p_X = p_Y \quad H_1: p_X \neq p_Y$

① Find $T_n(X_1, \dots, X_n, Y_1, \dots, Y_n)$

$\psi = \mathbb{1}\{T_n > s\}$

② Adjust s to guarantee asymptotic level α

③ type II error?

④ Remarks

Comparison of two proportions

① $\hat{p}_X = \frac{1}{n}\sum_{i=1}^n X_i \xrightarrow[n\to\infty]{\mathbb{P}} p_X$, $\quad \hat{p}_Y = \frac{1}{n}\sum_{i=1}^n Y_i \xrightarrow[n\to\infty]{\mathbb{P}} p_Y$

Consider $\hat{p}_X - \hat{p}_Y = g(\hat{p}_X, \hat{p}_Y)$, $\quad g(x,y) = x - y$

CLT: $$\sqrt{n}\left(\underbrace{\begin{pmatrix}\hat{p}_X\\ \hat{p}_Y\end{pmatrix}}_{=\frac{1}{n}\sum_{i=1}^n \begin{pmatrix}X_i\\ Y_i\end{pmatrix}} - \begin{pmatrix}p_X\\ p_Y\end{pmatrix}\right) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0, \Sigma)$$

$$\Sigma = \begin{pmatrix} p_X(1-p_X) & 0 \\ 0 & p_Y(1-p_Y)\end{pmatrix}$$

Delta Method:

$$\sqrt{n}\left(g(\hat{p}_X, \hat{p}_Y) - g(p_X, p_Y)\right) \to \mathcal{N}(0, \underbrace{\nabla g(p_X,p_Y)^T \Sigma \nabla g(p_X, p_Y)}_{\sigma_g^2})$$

$$\nabla g(x,y) = \begin{pmatrix}1\\ -1\end{pmatrix} \Rightarrow \boxed{\sigma_g^2 =} (1\ -1)\begin{pmatrix} p_X(1-p_X) & 0 \\ 0 & p_Y(1-p_Y)\end{pmatrix}\begin{pmatrix}1\\ -1\end{pmatrix}$$

$$\boxed{= p_X(1-p_X) + p_Y(1-p_Y)}$$

$X_1, \dots, X_n \sim \mathrm{Be}(p_X)$

$Y_1, \dots, Y_n \sim \mathrm{Be}(p_Y)$

iid. (Xs & Ys are independent)

$H_0: p_X = p_Y \quad H_1: p_X \neq p_Y$

① Find $T_n(X_1, \dots, X_n, Y_1, \dots, Y_n)$

$$T_n = \left| \sqrt{n} \frac{\hat{p}_X - \hat{p}_Y}{\sqrt{2\hat{p}(1-\hat{p})}} \right|$$

$$\psi = \mathbb{1}\{T_n > s\}$$

② Adjust s to guarantee asymptotic level α

③ type II error?

④ Remarks

Comparison of two proportions $\quad \hat{p}_X = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \hat{p}_Y = \frac{1}{n}\sum_{i=1}^{n} Y_i$

$$\sqrt{n}\left(\hat{p}_X - \hat{p}_Y - (p_X - p_Y)\right) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}\left(0, p_X(1-p_X) + p_Y(1-p_Y)\right)$$

$$\Rightarrow \sqrt{n} \cdot \frac{(\hat{p}_X - \hat{p}_Y - p_X + p_Y)}{\sqrt{p_X(1-p_X) + p_Y(1-p_Y)}} \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0,1)$$

For H_0, set $p_X = p_Y = p \in (0,1)$: $\quad p_X(1-p_X) + p_Y(1-p_Y) = 2p \cdot (1-p)$

$$\hat{p} = \frac{1}{2}(\hat{p}_X + \hat{p}_Y) \xrightarrow[n\to\infty]{\mathbb{P}} p$$

Slutsky's method: $\sqrt{n} \dfrac{\hat{p}_X - \hat{p}_Y}{\sqrt{2\hat{p}(1-\hat{p})}} \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0,1)$

$X_1, \dots, X_n \sim Be(p_X)$

$Y_1, \dots, Y_n \sim Be(p_Y)$

iid. (Xs & Ys are independent)

$p_X, p_Y \in (0,1)$

$H_0: p_X = p_Y \quad H_1: p_X \neq p_Y$

① Find $T_n(X_1, \dots, X_n, Y_1, \dots, Y_n)$

$$T_n = \left| \sqrt{n} \frac{\hat{p}_X - \hat{p}_Y}{\sqrt{2\hat{p}(1-\hat{p})}} \right|$$

$$\psi = \mathbb{1}\{T_n > s\}$$

② Adjust s to guarantee asymptotic level α

$s = q_{\alpha/2}$

③ Type II error

④ Remarks

Comparison of two proportions: $\hat{p}_X = \frac{1}{n}\sum_{i=1}^n X_i$, $\hat{p}_Y = \frac{1}{n}\sum_{i=1}^n Y_i$, $\hat{p} = \frac{1}{2}(\hat{p}_X + \hat{p}_Y)$

② $$T_n = \left| \sqrt{n} \frac{\hat{p}_X - \hat{p}_Y}{\sqrt{2\hat{p}(1-\hat{p})}} \right| \xrightarrow[n\to\infty]{\mathcal{D}} N(0,1) \text{ under } H_0 (p_X = p_Y)$$

$$\mathbb{P}(T_n > s) = \mathbb{P}\left(\left| \sqrt{n} \frac{\hat{p}_X - \hat{p}_Y}{\sqrt{2\hat{p}(1-\hat{p})}} \right| > s \right) \xrightarrow[n\to\infty]{} \mathbb{P}(|\underbrace{Z}_{\sim N(0,1)}| > s) \stackrel{!}{=} \alpha$$

$$= 2 \cdot (1 - \Phi(s))$$

$$\Rightarrow s = q_{\alpha/2}, \ 1-\alpha/2 \text{ quantile of } N(0,1)$$

③ $$\hat{p}_X \xrightarrow[n\to\infty]{\mathbb{P}} p_X, \ \hat{p}_Y \xrightarrow[n\to\infty]{\mathbb{P}} p_Y, \ \hat{p} = \frac{1}{2}(\hat{p}_X + \hat{p}_Y) \xrightarrow[n\to\infty]{\mathbb{P}} \frac{1}{2}(p_X + p_Y) =: \tilde{p}$$

$$T_n = \left| \sqrt{n} \frac{\overbrace{\hat{p}_X - \hat{p}_Y}^{\to\, p_X - p_Y \neq 0 \text{ under } H_1}}{\underbrace{\sqrt{2\hat{p}(1-\hat{p})}}_{\to \sqrt{2\tilde{p}(1-\tilde{p})}}} \right| \xrightarrow[n\to\infty]{\mathbb{P}} +\infty \Rightarrow \text{type II error} \xrightarrow[n\to\infty]{\mathbb{P}} 0$$

$X_1,\dots,X_n \sim Be(p_X)$

$Y_1,\dots,Y_n \sim Be(p_Y)$

iid. (Xs & Ys are independent)

$p_X, p_Y \in (0,1)$

$H_0: p_X = p_Y$ $\quad H_1: p_X \neq p_Y$

① Find $T_n(X_1,\dots,X_n,Y_1,\dots,Y_n)$

$$T_n = \left|\sqrt{n}\,\frac{\hat{p}_X - \hat{p}_Y}{\sqrt{2\hat{p}(1-\hat{p})}}\right|$$

$$\psi = \mathbb{1}\{T_n > s\}$$

② Adjust s to guarantee asymptotic level α

$s = q_{\alpha/2}$

③ Type II error

④ Remarks

Comparison of two proportions

④ Different sample sizes: $X_1,\dots,X_{n_1},\ Y_1,\dots,Y_{n_2}$

CLT: $Z_1,\dots,Z_n \in \mathbb{R}$ iid, $\mathbb{E}[Z_i]=\mu$, $Var(Z_i)=\sigma^2$

$$\underbrace{\sqrt{n}\,\frac{\bar{Z}_n-\mu}{\sqrt{\sigma^2}}}\ \xrightarrow[n\to\infty]{D} N(0,1);\quad \bar{Z}_n = \frac{1}{n}\sum_{i=1}^{n} Z_i$$

$\mathbb{E}[-n-]=0$, $Var(-n-)=1$

$$\boxed{\mathbb{E}[\bar{Z}_n] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[Z_i] = \frac{n}{n}\mu = \mu}$$

$$\boxed{Var(\bar{Z}_n) = \frac{1}{n^2}\sum_{i=1}^{n} Var(Z_i) = \frac{n}{n^2}Var(Z_i) = \frac{1}{n}\sigma^2}$$

Consider $\hat{p}_X - \hat{p}_Y$,

$\hat{p}_X = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$, $\hat{p}_Y = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i$. $Var(\hat{p}_X - \hat{p}_Y) = ?$ under H_0